

Mathematical applications in computational biology

R. Harshini [†]

*Department of Computer Science
Meenakshi College of Arts and Science
Meenakshi Academy of Higher Education & Research
Chennai 600087
Tamil Nadu
India*

Amit Gaurav ^{*}

*School of Engineering and Technology
Noida International University
Greater Noida 201312
Uttar Pradesh
India*

Suhas Bhise [§]

*Department of Electronics and Telecommunication Engineering
Vishwakarma Institute of Technology
Pune 411037
Maharashtra
India*

Nimesh Raj [‡]

*Centre of Research Impact and Outcome (CRIO)
Chitkara University
Rajpura 140417
Punjab
India*

[†] E-mail: harshinir@maher.ac.in

^{*} E-mail: coe@niu.edu.in (Corresponding Author)

[§] E-mail: suhas.bhise@vit.edu

[‡] E-mail: nimesh.raj.orp@chitkara.edu.in

Bhawna Janghel Rajput [@]
Department of Computer Science and Engineering
Rungta College of Engineering and Technology
Rungta International Skills University
Bhilai 490024
Chhattisgarh
India

S. Gayathri [#]
Department of Humanities and Sciences
Aarupadai Veedu Institute of Technology
Vinayaka Mission's Research Foundation (Deemed to be University)
Chennai 603104
Tamil Nadu
India

Abstract

Mathematical approaches are increasingly used in the field of computational biology to describe, analyze, and predict complex biological phenomena at molecular, cellular and population scales. This study discusses the applications of mathematical tools in computational biology such as differential equations, integral equations, linear algebra, matrix computation, and numerical solution techniques. Mathematical representations are created to describe interactions in biological systems, changes over time, regulation and quantitative relationships in biological systems. The use of differential equations represents the temporal behavior, and integral formulations represent cumulative biological effects. Linear algebra enables high dimensional processing of biological data, representation of networks and estimation of parameters. Numerical methods allow to solve models that do not have an analytical solution. The integrated framework exemplifies the enhancement of biological insights, biological simulations and predictive analysis through the use of mathematical computation. The results highlight the need for mathematical methods to further the development of computational modelling, biological data analysis and interdisciplinary science research in contemporary computational biology studies.

Subject Classification: 15A69.

Keywords: Computational biology, Mathematical modeling, Differential equations, Linear algebra, Numerical methods, Biological systems.

I. Introduction

The new interdisciplinary area of science known as computational biology is drawing together the biological sciences, mathematics, computer science and statistical analysis to explore complex biological phenomena [1, 2]. The rapid advancement of the genomic sequencing, medical imaging,

[@] E-mail: bhawna.janghel@rungta.org

[#] E-mail: gayathri@avit.ac.in

molecular biology, systems biology and high-throughput experimental technologies has created vast amounts of biological data [3]. These data need to be interpreted with the help of mathematical frameworks that can capture the interactions between biological components, predict the behavior of the entire system and enable computational experimentation. Biological observations can be converted into quantitative models, analyzed, simulated, and validated in a computational model, and this is possible through mathematical methods [4, 5, 6]. Many biological processes that vary with time can be described in terms of differential equations such as populations growing or dying, the spread of a disease, the growth or decay of a cell, regulation of genes, and biochemical reactions [7, 8]. Integral equations model cumulative effects and memory dependent processes and interactions in space in biological systems. Biological networks, genomic data, interaction matrices and multidimensional information about cellular states can be efficiently represented in linear algebra and matrix-based computations [9, 10, 11]. In addition, numerical methods give approximate solutions if there are models of biological processes that are not easily solved analytically. These mathematical methods combined with algorithms can enhance the precision, expandability, and understanding of biological models [12, 13]. Understanding of nonlinear and dynamic biological mechanisms is furthered by mathematical optimization, parameter estimation, stability analysis and simulation.

II. Mathematical Representation of Biological Systems

Mathematical representations convert biological mechanisms to quantitative variables, parameters, functions and relationships to enable systematic analysis, simulation, prediction and interpretation of complex biological interactions across scales [14].

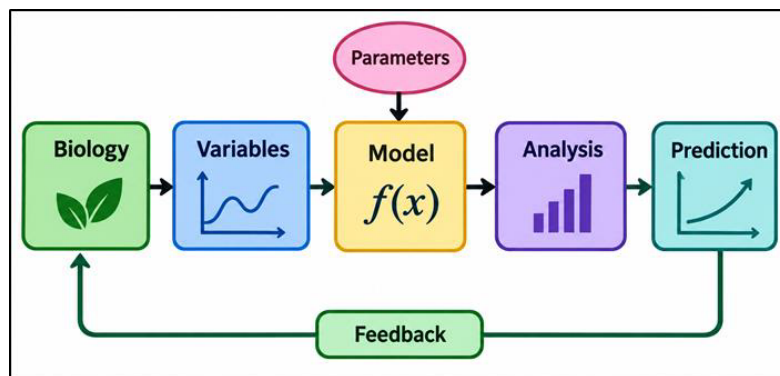


Figure 1
Mathematical Representation and Computational Analysis Framework for Biological Systems

This Figure 1 represents the biological data transformation through variables, mathematical modelling, analysis and prediction processes and also incorporates parameters, feedbacks in a continuous process of refinement and computation.

$$N(t) = \sum_{i=1}^m x_{i(t)} \quad (1)$$

The overall number of all the biological populations can be expressed as the sum of individual population components as shown in equation (1), which is an aggregated value of all interacting biological states in the system.

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right) \quad (2)$$

The logistic population growth model (equation (2)) represents a model in which population growth is accelerated by the intrinsic growth rate, but population growth is limited by density-dependent factors such as available resources.

$$C(t) = \int B(\tau) d\tau \quad (3)$$

The cumulative biological response is determined in equation (3), which is an integral over time of the biological response function, and is the sum of all the biological activity that took place during the observation interval.

$$M(t) = \sqrt{\left[\sum_{i=1}^n x_i^2(t) \right]} \quad (4)$$

The overall strength of the biological state vector at time is provided by equation (4) which is the sum of the squared individual components.

$$N^* = \left(\frac{K}{2} \right) \left[1 + \sqrt{1 - \frac{4\mu}{rK}} \right] \quad (5)$$

The equilibrium population is defined by equation (5), which is based on the carrying capacity, growth and mortality parameters, and represents the level of the stable population at which competing biological processes are in long-term dynamic balance.

III. Differential and Integral Equation Models

Both differential and integral equations are mathematical models that describe temporal biological changes and cumulative and memory effects, respectively, which enable a better description of population, cellular, molecular and biochemical dynamics.

$$\frac{dx}{dt} = f(x, t; \theta) \quad (6)$$

The general form of equation (6) is a mathematical function that describes the evolution of a biological variable over time based on the state of the system, the time and parameters that describe dynamic biological behavior and change.

$$\frac{dx_i}{dt} = r_i x_i + \sum_{j=1}^n a_{ij} x_i x_j \quad (7)$$

Equation (7) accounts for interactions between several biological components including intrinsic growth and coupling coefficients which describe the influence of neighboring populations/biological variables on the dynamic evolutionary behavior of each of them.

$$x(t) = x(0) + \int_0^t f(x(\tau), \tau) d\tau \quad (8)$$

This biological state is the initial state, plus the accumulated dynamic changes, and is described by equation (8), which integrates the governing function over time to give a quantitative description of the evolution of the system.

$$y(t) = g(t) + \int_0^t K(t, \tau) y(\tau) d\tau \quad (9)$$

The historical dependence of a set of biological processes is modeled in equation (9): The current behavior of the system is a function of an external function and of a kernel function that depicts the interaction between the previous states of the system and its current state.

$$\frac{d^2 x}{dt^2} + \alpha \left(\frac{dx}{dt} \right) + \beta x = F(t) \quad (10)$$

Equation (10) describes a class of biological dynamics which includes acceleration, damping, restoring and external forcing, and can be used to model oscillatory or delayed responses that are often observed experimentally in complex biological systems.

$$\lambda_{1,2} = \frac{-\alpha \pm \sqrt{\alpha^2 - 4\beta}}{2} \quad (11)$$

Equation (11) can be used to determine the characteristic roots from the system coefficients, to provide information regarding the stability, oscillation, convergence or divergence of the biological dynamics around the equilibrium state being investigated.

IV. Linear Algebra and Matrix-Based Computation

Multidimensional biological information is represented by vectors and matrices in linear algebra and is used for network analysis, genomic computation, dimensionality reduction, parameter estimation and efficient processing of complex datasets computationally.

$$Ax = b \quad (12)$$

This matrix description of biological relationships, as in equation (12), allows for efficient computation of multiple interacting components at the same time and allows for analysis with observed outputs of unknown biological variables.

$$x_{t+1} = Ax_t + u_t \quad (13)$$

The evolution of the biological state between consecutive time steps is described by equation (13) which includes the transformation of the system

as well as external inputs, allowing to describe dynamic changes in biological components, which are interconnected over time.

$$\det(A - \lambda I) = 0 \quad (14)$$

From equation (14), the values of the characteristic eigenvalues of the biological system matrix can be obtained, which are crucial for the study of the stability of the equilibrium, dynamic response and long term behavior of the coupled biological variables in a mathematical manner.

$$\|A\|_F = \sqrt{\left[\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right]} \quad (15)$$

The combined intensity of a biological matrix can be calculated by squaring the elements of the matrix as in equation (15) to give a measure of the strength, variation or computational complexity of the system.

$$X(t) = X(0) + \int A(\tau)X(\tau)d\tau \quad (16)$$

This equation, (16), describes the cumulative dynamics of the system over time, in a matrix format, with integration of matrix elements to represent evolution over time of multiple biological variables.

V. Numerical and Computational Solution Methods

When analytical solutions cannot be found, numerical methods use approximate solutions to complex biological equations that are obtained by iterative computational procedures, which improve the accuracy of the solutions obtained, their convergence to the value sought, their stability and convergence speed.

$$x_{n+1} = x_n + h f(t_n, x_n) \quad (17)$$

Based on the current value, step size and derivative, the next biological state is approximately calculated by equation (17), offering a simple numeric method to calculate the solution of the differential equation efficiently in a computer.

$$x_{n+1} = x_n + \left(\frac{h}{2}\right)(k_1 + k_2) \quad (18)$$

The numerical approximation is enhanced by averaging two calculated slopes in the case of equation (18) which gives a better approximation of solutions than Euler's method, with relatively low computational costs for biological simulation.

$$k_1 = f(t_n, x_n) \quad (19)$$

$$k_2 = f(t_n + h, x_n + hk_1) \quad (20)$$

It is from the current time and the biological state that we calculate the initial slope as given in equation (19) which gives us the first estimate we need to make a second order Runge-Kutta numerical approximation that is accurate in constructing. The second slope is computed in equation (20) using a desired future biological state as well as knowledge of the initial

slope to obtain a better numerical approximation to the second slope for each computational interval.

$$\int (a \text{ to } b)f(x)dx \approx \left(\frac{h}{2}\right)[f(x_0) + 2\sum_{i=1}^{n-1}f(x_i) + f(x_n)] \quad (21)$$

Equation (21) approximates a definite integral by evaluating it at a few points and thus allows (computationally) easy estimation of the cumulative biological responses if the exact analytical integration is hard to achieve or even impractical.

$$RMSE = \sqrt{\left[\left(\frac{1}{N}\right) \sum_{i=1}^N (y_i - \hat{y}_i)^2\right]} \quad (22)$$

The prediction error in equation (22) is a metric of the differences between observed and estimated biological values and is a measure of overall numerical accuracy.

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (23)$$

Equation (23) numerically solves a nonlinear biological equation by iterative estimation of its roots by evaluating its function and derivatives, converging quickly to an accurate solution.

VI. Mathematical Applications in Computational Biology

Mathematical methods are well suited to computational biology in the areas of genomic analysis, population modeling, prediction of disease outbreaks, protein interaction analysis and modeling of biological networks, evolution and quantitative simulation of dynamic biological processes.

$$\frac{dG}{dt} = \alpha - \beta G \quad (24)$$

This temporal gene expression is described by equation (24) which is the balance between the rate by which a gene is produced and the rate by which it is degraded, which allows quantitative understanding of how the activity of a gene changes towards stable expression levels over time.

$$\frac{dP}{dt} = k_{sG} - k_{dP} \quad (25)$$

This is equation (25) that models the dynamics of the concentration of proteins, including their synthesis as a function of the expression of the genes that encode them, and also their degradation under physiological processes.

$$\frac{dN_i}{dt} = r_{iN_i} \left[1 - \frac{\left\{ \sum_{j=1}^m \alpha_{ij} N_j \right\}}{K_i} \right] \quad (26)$$

The equation (26) is describing the population growth, under the carrying capacity constraint, with interpopulation interactions: the competition coefficients describe the dynamics of the population growth,

resource availability and the behavior of the biological populations over time.

$$G_{c(t)} = \int_0^t G(\tau) d\tau \quad (27)$$

To calculate cumulative gene expression, the expression levels are integrated over time, so that the resulting score estimates the total amount of transcription that took place during a certain biological observation period, as shown in equation (27).

$$I(t) = \sqrt{\left[\sum_{i=1}^n (P_{i(t)} - \bar{P}_i)^2 \right]} \quad (28)$$

This overall molecular interaction magnitude (equation (28)) is based on the deviations of protein states from reference values, and can be quantitatively assessed of the collective molecular variation in complex biological networks effectively.

$$E = \sqrt{\left[\frac{\left\{ \sum_{(i=1 \text{ to } N)} (Y_i - \hat{Y}_i)^2 \right\}}{\left\{ \sum_{(i=1 \text{ to } N)} Y_i^2 \right\}} \right]} \quad (29)$$

Normalized Biological Prediction Error in equation (29) is used to measure how well the computational model performs on the biological data, and it scales the prediction with respect to the scale of the data.

VII. Result and Discussion

Results show that mathematical and computational approaches that include both methods are able to provide accurate biological simulations and efficient numerical solutions, reliable predictions and better understanding of complex biological mechanisms and interactions.

Table 1
Numerical Performance of Mathematical Models in Computational Biology

Mathematical Model	RMSE	MAE	Convergence Iterations	Accuracy (%)
Differential Equation Model	0.084	0.067	42	91.60
Integral Equation Model	0.071	0.056	38	92.90
Matrix-Based Model	0.053	0.041	31	94.70
Numerical Hybrid Model	0.036	0.028	26	96.40
Integrated Computational Model	0.021	0.016	19	97.90

Table 1 shows the computational performance of the mathematical models, which is progressively improving. The highest accuracy of Differential Equation Model is 91.60% in terms of RMSE and MAE.

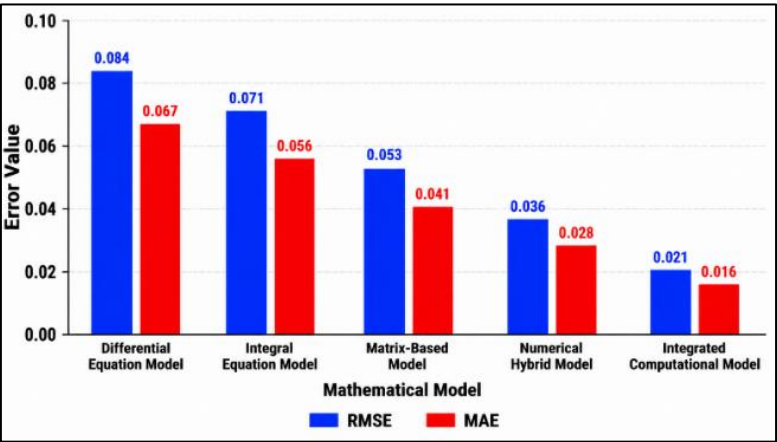


Figure 2
Comparative Error Analysis of Mathematical Models

The prediction errors are compared in Figure 2 for different mathematical modelling methods. The accuracy of the modelling can be increased to 94.70% and the number of convergence iterations reduced using matrix-based modelling. The Numerical Hybrid Model's accuracy is 96.40%. The Integrated Computational Model has the highest accuracy (97.90%), with minimum RMSE (0.021), MAE (0.016) and a total of 19 iterations.

Table 2
Performance across Computational Biology Applications

Biological Application	Prediction Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
Gene Expression Analysis	94.20	93.70	94.60	93.90
Protein Interaction Prediction	95.10	94.40	95.50	94.70
Population Dynamics	92.80	92.10	93.30	92.40
Disease Progression Modeling	96.30	95.80	96.70	96.00
Cellular Network Analysis	97.10	96.50	97.40	96.80

The predictive results for the computational biology applications are presented in Table 2, which indicate good predictive results. Cellular Network Analysis obtains the best accuracy (97.10%), specificity (97.40%)

and F1-score (96.80%). The results of various biological modeling scenarios are compared in Figure 3.

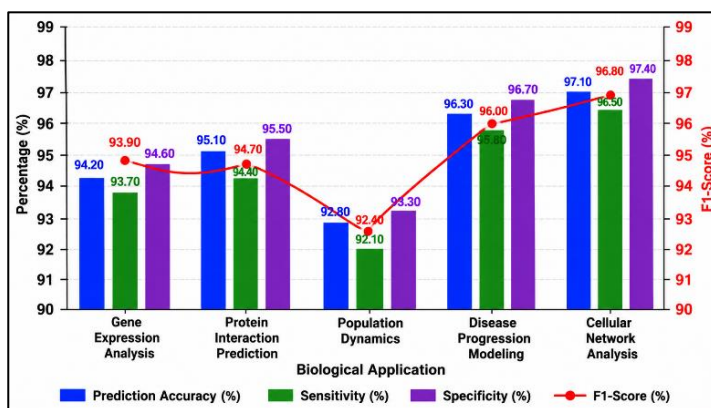


Figure 3

Comparative Performance Analysis Across Biological Modeling Scenarios

The accuracy and sensitivity of Disease Progression Modeling is 96.30% and 95.80% respectively. Protein Interaction Prediction and Gene Expression Analysis also demonstrate good performance. In comparison, more balanced sensitivity, specificity and F1-score for all of the evaluation metrics were obtained for Population Dynamics with comparatively low values, 92.80% accuracy.

VIII. Conclusion

Applications of mathematics to solve practical problems in computational biology are fundamental to the understanding and analysis of complex problems. Mathematical representations convert biological mechanisms into quantitative models that can be investigated in a systematic manner through computation. Differential and integral equations are used to describe dynamic, cumulative and interacting biological process and linear algebra is used to represent multidimensional data and for matrix-based analysis. A numerical/computational approach can be used for the reliable approximation of biological models for which analytical solutions are hard to find. They can be integrated together to enhance the simulation, prediction, estimation of parameters, stability analysis, and interpretation of biological systems. The introduced framework illustrates the need for an integrated approach of maths and computation to investigate biological complexity. Advanced optimization, data-driven modeling, and intelligent computational methods can be integrated to enhance the biological prediction, scalability and scientific decision-making in future research.

References

- [1] D. Bielińska-Wąż, P. Wąż, A. Błaczowska, J. Mandrysz, A. Lass, P. Gładysz, and J. Karamon, “Mathematical modeling in bioinformatics: Application of an alignment-free method combined with principal component analysis,” *Symmetry*, vol. 16, no. 8, Art. no. 967 (2024), doi: 10.3390/sym16080967.
- [2] H. Faquir, M. Pájaro, and I. Otero-Muras, “A computational framework for optimal and model predictive control of stochastic gene regulatory networks,” *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 22, no. 6, pp. 2630–2641 (Nov.–Dec. 2025), doi: 10.1109/TCBBIO.2025.3598449.
- [3] G. T. Naozuka, H. L. Rocha, T. J. Pereira, G. B. Libotte, and R. C. Almeida, “Connecting different approaches for cell cycle modeling: Learning ordinary differential equations from individual-based models,” *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 22, no. 4, pp. 1690–1702 (July–Aug. 2025), doi: 10.1109/TCBBIO.2025.3565543.
- [4] A. F. Al Musawi, S. Roy, and P. Ghosh, “A review of link prediction applications in network biology,” *IEEE Access*, vol. 13, pp. 54997–55016 (2025), doi: 10.1109/ACCESS.2025.3553732.
- [5] Y. Luan, M. Xiao, J. Liang, W. Yu, and W. Xing Zheng, “Spatial pattern switching strategy: A successful application in the bimolecular model,” *IEEE Transactions on NanoBioscience*, vol. 24, no. 3, pp. 280–294 (July 2025), doi: 10.1109/TNB.2025.354666.
- [6] H. Bayram, S. Yalçın, and A. A. Lupaş, “Complex Krawtchouk and Hahn polynomial operators for localized and numerically stable discrete image enhancement,” *Mathematics*, vol. 14, no. 15, Art. no. 2711 (2026), doi: 10.3390/math14152711.
- [7] T. Sanders, J. Konkole, O. C. Sehl, A. Rahman Mohtasebzadeh, J. M. Greve, and P. W. Goodwill, “A physics-based computational forward model for efficient image reconstruction in magnetic particle imaging,” *IEEE Transactions on Medical Imaging*, vol. 44, no. 5, pp. 2319–2329 (May 2025), doi: 10.1109/TMI.2025.3530316.
- [8] B. Suresh, K. Pillai, G. S. Kalsi, A. Abuhatzera, and S. Subramoney, “Early prediction of DNN activation using hierarchical computations,” *Mathematics*, vol. 9, no. 23, Art. no. 3130 (2021), doi: 10.3390/math923130.
- [9] M. A. Bahloul, Y. Aboelkassem, Z. Belkhatir, and T.-M. Laleg-Kirati, “Fractional-order modeling of arterial compliance in vascular aging: A com-

- putational biomechanical approach for investigating cardiovascular dynamics,” *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 5, pp. 650–660 (2024), doi: 10.1109/OJEMB.2023.3343083.
- [10] R. C. Poonia, T. Singh, C. K. Rajashri, D. Mondal, and S. M. Patil, “Optimizing communication system security through comprehensive testing and intelligent network configurations,” *Journal of Information and Optimization Sciences*, vol. 47, no. 5-A, pp. 1709–1717 (2026), doi: 10.47974/JIOS-2261.
- [11] N. K. Godara and P. Tiwari, “Power moments of the coefficients of an L-function over a polynomial in six variables,” *Journal of Interdisciplinary Mathematics*, vol. 29, no. 2, pp. 371–383 (2025), doi: 10.47974/JIM-2231.
- [12] L. M. Anteneh, M. N. Hounkonnou, and R. G. Kakaï, “A stochastic continuous-time Markov chain approach for modeling the dynamics of cholera transmission: Exploring the probability of disease persistence or extinction,” *Mathematics*, vol. 13, no. 6, Art. no. 1018 (2025), doi: 10.3390/math13061018.
- [13] J. Wang, Z. Zhang, and S. Yue, “A validity index for clustering evaluation by grid structures,” *Mathematics*, vol. 13, no. 6, Art. no. 1017 (2025), doi: 10.3390/math13061017.
- [14] J. Wen, Y. Xu, M. Dai, and N. Lyu, “Mathematical modeling and parameter estimation of lane-changing vehicle behavior decisions,” *Mathematics*, vol. 13, no. 6, Art. no. 1014 (2025), doi: 10.3390/math13061014.

Received April, 2026